

# Обучение машинного перевода без параллельных текстов\*

*Рябов Ф. А. Бахтеев<sup>1</sup> О. Ю. Стрижов<sup>2</sup> В. В.*

<sup>1</sup>Московский физико-технический институт

<sup>2</sup>Вычислительный центр им. А. А. Дородницына ФИЦ ИУ РАН

В работе рассматривается задача построения системы машинного перевода. В стандартном подходе к такой задаче используется набор параллельных предложений, состоящих из пар предложений, являющихся переводом друг друга. Однако в таком случае система нуждается в огромном объеме тестовой выборки, исчисляемой миллионами. В данной задаче рассматривается метод построения модели машинного перевода без использования параллельного корпуса. Основой метода является Seq2seq, где предложениям каждого языка ставится в соответствие вектор из общего векторного пространства. Вычислительный эксперимент проводится на паре "русский-украинский".

**Ключевые слова:** *машинный перевод, нейросеть, Seq2seq.*

---

\*Работа выполнена при финансовой поддержке РФФИ, проект №00-00-00000. Научный руководитель: Стрижов В. В. Авторы: А. В. Грабовой, О. Ю. Бахтеев, В. В. Стрижов, Eric Gaussier, координатор Малиновский Г. С. Консультант: Бахтеев О. Ю.